

Response to: “The Worldwide Governance Indicators: Six, One, or None”

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi¹

January 2010

While in graduate school, one of the authors of this note worked as a teaching assistant for an undergraduate introductory statistics course. One TA session was devoted to a discussion of the distinction between causation and correlation. A favourite example was the strong positive correlation between university education and subsequent earnings observed across individuals. Students invariably quickly advanced two hypotheses to explain this correlation: (1) higher education created useful skills which were subsequently rewarded by higher earnings, and (2) “smart” kids got into university and also naturally earned more later in life because they were “smart”. For some reason (hopefully not reflecting the quality of instruction in that particular TA session!), students tended to be more sympathetic to the second hypothesis than the first.

Had Steve Knack and Laura Langbein (KL) attended this particular TA session the discussion might have taken an unexpected turn. Reasonably enough KL would have agreed that these are interesting hypotheses and would have suggested empirically testing them. As a test of the first hypothesis, they would have proposed estimating an Ordinary Least Squares (OLS) regression of earnings on education, and would have observed a high R-squared. As a test of the second hypothesis, they would have proposed measuring the correlation between education and earnings, and would again have observed that it was high. Then KL would have shared with the class their shocking (to them) conclusion that the data strongly supported both hypotheses, since the two tests led to exactly the same results. At this point the TA might have interrupted the discussion to gently point out that finding identical results from the two tests was not so surprising since after all the R-squared from the first regression was just the square of the correlation between education and earnings. The TA might also have pointed out that the assumption of an exogenous right-hand-side variable required to justify KL’s OLS regression was very unlikely to be valid in this case, precisely because of the possibility of an omitted variable such as “smarts”. Despite this, KL would have continued undaunted to their ultimate conclusion: in an impressive *non sequitur*, they would have argued that since the data could not discriminate between these two hypotheses, the data itself must constitute a “tautology” and that the empirical measures of education and earnings were useless precisely because they were so correlated with each other.

Readers should note that the first paragraph is fact, and the second is fiction, but just barely, since it captures the essence of the peculiar logic contained in Knack and Langbein’s (2010) critique of

¹ Brookings Institution, World Bank Development Research Group, and World Bank Institute, respectively. The authors can be contacted at dkaufmann@brookings.edu, akraay@worldbank.org, and mmastruzzi@worldbank.org. The views expressed here do not reflect those of the Brookings Institution, The World Bank, its Executive Directors, or the countries they represent.

the Worldwide Governance Indicators (WGI). In the WGI project we combine data from a large number of underlying sources to construct six aggregate governance indicators: Voice and Accountability (VA), Political Stability and Absence of Violence (PV), Government Effectiveness (GE), Regulatory Quality (RQ), Rule of Law (RL), and Control of Corruption (CC). The indicators cover over 200 countries over the period 1996-2008, and as KL note, the WGI are widely used in scholarly research, and in development policy discussions.²

After recycling a variety of pre-existing purported critiques of the WGI, KL move to the core contribution of their paper.³ KL reasonably posit that there might be a variety of causal relationships between the dimensions of governance we seek to measure in the WGI. For example, one might sensibly hypothesize that an absence of democratic accountability (VA) might foster corruption (CC). In order to test this “causal model” they propose estimating an OLS regression of CC on VA. They also include in this regression the four other WGI measures as right-hand-side variables. However this is inessential to our illustration of their argument, and so for simplicity we focus only on a minimal version of their “causal model” in which VA alone “causes” CC. KL estimate this model by OLS and observe that it fits the data very well, which of course is nothing more than the observation that the correlation between VA and CC is large.

KL next propose testing a “measurement model” in which all six of the WGI measures are driven by some unobserved common factor that KL refer to as “good government”. To test this model they use factor analysis to extract the first principal component of the six WGI measures, and observe that it accounts for much of the variation in the WGI. They interpret this as strong evidence in favour of the “measurement model”. Again, for our purposes it is irrelevant that KL use all six WGI measures in this exercise, and so to simplify the discussion, imagine instead using factor analysis to extract a common factor from just two of the WGI variables, VA and CC. The share of the variation in VA and CC “explained” by the common factor is nothing more than the correlation between VA and CC. And so it is entirely unsurprising that KL find that both the “causal” and the “measurement” models fit the data equally well – it is because the tests of both models are based on precisely the same statistic: the correlation between VA and CC, which happens to be large.

If we replace “education” with “VA” and earnings with “CC” we can see how the story of the introductory statistics TA session captures the essence of KL’s paper. Just as education and earnings are correlated in the data, so are VA and CC. This correlation might be due to a causal impact from the one to the other (in either direction), or it may reflect the effect of some unobserved confounding factor, “smarts” in the case of education and earnings, or “good government” in the case of VA and CC. Simply observing this correlation tells us nothing about the mechanism that generates it. In other words, these hypotheses are *observationally equivalent* as both imply a strong positive correlation between

² The WGI are described in a series of papers, the latest of which is Kaufmann, Kraay and Mastruzzi (2009). According to Google Scholar the WGI papers have been cited in over 3500 scholarly papers over the past decade.

³ We do not respond to all these other critiques here as none of them are new, and have been refuted as either incorrect or unsubstantiated speculation in our earlier work, see especially Kaufmann, Kraay and Mastruzzi (2007a, 2007b, and 2010).

education and earnings, or VA and CC. Absent further identifying information, which KL singularly fail to provide, the underlying mechanisms responsible for the correlation cannot be isolated.

That correlation does not identify causation is hardly news: it has after all been nearly a century since pioneering econometricians such as Wassily Leontieff in the 1920s recognized that the correlation between prices and quantities is uninformative about the slopes of supply and demand curves. And a key feature of empirical work in economics over the past several decades is the quest for better identification strategies that permit the isolation of causal effects. These have included greater reliance on randomization, and an increasingly-creative quest for natural experiments that provide plausibly exogenous variation in explanatory variables of interest. Against this background, KL's specification and estimation of their "causal model" is breathtakingly naive. They describe their OLS regression which they claim estimates the "causal model" as follows: *"In this model CC is endogenous, and the other WGI variables are exogenous, as is an error term for CC. The error term and the exogenous WGI variables are uncorrelated..."* (KL 2010, p. 361). Such a claim is indefensible, not least because it contradicts KL's own subsequent hypothesis that there might be an unobserved common factor driving all dimensions of governance, or that there might equally well be causation in the opposite direction.⁴ Either of these would of course induce a correlation between the error term and the regressors that renders OLS invalid.

In the end, KL's "causal" and "measurement" models are nothing more than two different ways of documenting exactly the same empirical fact: the six WGI measures are strongly correlated among themselves. One does not need to read much of KL's paper to recognize this fact, as it is already documented directly in their own Table 1. Since KL provide no credible identifying assumptions of any sort, we learn nothing from the empirical work that follows beyond what is already captured by this simple table of correlations. And in particular, we learn nothing about whether the "causal" or "measurement" models are responsible for generating the observed correlation: they are after all observationally equivalent, despite KL's claims to the contrary as they try to estimate both models.

The conclusion is the most peculiar part of KL's paper. Based on their finding that the "causal" and "measurement" models both fit the data equally well, they leap with no further justification to the conclusion that the WGI somehow represent a "tautology", and that the WGI fail to measure what they claim to measure. If we were to take KL's peculiar logic seriously, we would have to conclude that *any* two variables that are correlated with each other would be invalid empirical measures, since they would provide what KL refer to as strong evidence for both a "causal" and a "measurement" model linking the two variables. And even if one were to provide properly-identified evidence for a causal effect of one variable on another, or for an omitted variable driving both, KL provide no coherent reason for thinking that this somehow reflects a problem in the data itself. This would be like concluding that data on years of schooling and earnings somehow are not valid measures of education and income simply because there is a correlation between the two. Such an unsubstantiated conclusion in the context of education

⁴ For example, there is a large literature suggesting that corruption undermines confidence in, and the functioning of, democratic institutions. See Clausen, Kraay, and Nyiri (2009) for a recent contribution, and a thorough discussion of the identification problem in that context.

and earnings would have earned a failing grade in the introductory statistics TA session described above. It fares no better in the context of the WGI.

References:

Clausen, Bianca, Aart Kraay and Zsolt Nyiri (2009). "Corruption and Confidence in Public Institutions: Evidence from a Global Survey". World Bank Policy Research Department Working Paper No. 5157.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2007). "The Worldwide Governance Indicators: Answering the Critics". World Bank Policy Research Department Working Paper No. 4149.

_____, (2007b). "Growth and Governance: A Reply/Rejoinder". *Journal of Politics*. 69(2):555-562, 570-572.

_____, (2009). "Governance Matters: Aggregate and Individual Governance Indicators 1996-2008". World Bank Policy Research Department Working Paper No. 4978.

_____, (2010). "Response to: What Do the Worldwide Governance Indicators Measure?". *European Journal of Development Research*, forthcoming

Knack, Stephen and Laura Langbein (2010). "The Worldwide Governance Indicators: Six, One, or None?". *Journal of Development Studies*. 46(2):350-370.